ELSEVIER

Reliability Engineering and System Safety 64 (1999) 365–381

RELIABILITY ENGINEERING & SYSTEM SAFETY

Trending of equipment inoperability for commercial aircraft

James T. Luxhøj*

Department of Industrial Engineering, Rutgers, The State University of New Jersey, 96 Frelinghuysen Road, Piscataway, NJ 08854-8018, USA

Received 22 May 1998; accepted 5 September 1998

Abstract

The Safety Performance Analysis System (SPAS) is a Web-based analytical tool, or rather a database system, that is being fielded by the Federal Aviation Administration (FAA). It is intended to provide the FAA's Aviation Safety Inspectors (ASIs) with the means to evaluate and control appropriate surveillance levels for aircraft operators. The SPAS enables qualitative trend analysis to be performed at the national, regional, and district levels. This will permit the system to be more sensitive to particular areas of concern for a given region. For example, inspectors particularly desire a tool to help them recognize problems which are occurring in the fleet. Especially desirable is a tool that could do so automatically by searching for trends and alerting the relevant inspectors.

The Service Difficulty Reporting (SDR) system, one of the SPAS databases, provides FAA inspectors with information related to aircraft equipment inoperability, such as in-service difficulties, malfunctions, and defects. The SDR data provide information for planning, directing, controlling, and evaluating certain assigned safety and maintenance programs. This paper presents previous SDR trending results for DC-9 operators with heterogeneous fleets, but now uses data from two specific operators with homogenous fleets consisting entirely of 737 aircraft. This new research is an extension of the previous SDR forecasting research, but now provides more meaningful information, as the DC-9 aircraft data were composed of numerous operators with mixed fleets and differing operating and maintenance policies. Multiple regression and neural network models are the principal two forecasting methods examined. A population modeling concept, or data grouping strategy, appears to be an effective technique for trending SDRs for operators of either heterogeneous and homogeneous fleets. The forecasting methods presented in this paper offer technical enhancements for SDR trending compared to the current qualitative method of visual observation of graphical plots.

Keywords: Safety performance analysis system; Equipment trending; Population modeling; Data grouping strategy

1. Introduction

Inspection and diagnostic activities are integral components of an effective maintenance strategy in an attempt to ensure aviation system safety, reliability, and availability. The Federal Aviation Administration (FAA) in the United States is responsible for regulating aircraft traffic and safety. An expected increase in usage of domestic flights in the next few years coupled with an aging population of aircraft has led the FAA to initiate new aviation safety research efforts [1]. The Safety Performance Analysis System (SPAS) is an analytical tool, or rather a database system, that is currently in a production version [2-4]. The SPAS steering committee was established in 1991, the first prototype was launched in 1993 and operational testing was completed in 1995. The production version, SPAS II, is a Web-based software that is intended to provide the FAA's Aviation Safety Inspectors (ASIs) with the means to evaluate and control appropriate surveillance levels for aircraft operators. As of July 1998, approximately 50% of the estimated 3400 inspector workforce from various FAA Fight Standards District Offices (FSDOs) across the United States have already completed software training.

The SPAS integrates a number of existing FAA databases, and enables data analysis to be performed at the national, regional, and district levels. This will permit the system to be more sensitive to particular areas of concern for a given region. For example, inspectors particularly desire a tool to help them recognize problems which are occurring in the fleet. Especially desirable is a tool that could do so automatically by searching for trends and alerting the relevant inspectors.

The SPAS also supports the user's ability to check for the occurrence of problems identified as related, i.e. it identifies patterns. This includes the occurrence of patterns which are cyclic or repeated. It includes patterns which result from comparisons to group or industry standards, or which occur in a clustered fashion. Patterns emerging from the data will be able to trigger an alert for the inspector who

* Tel.: + 1-732-445-3625; Fax: + 1-732-445-5467; e-mail: luxhoj@princess.rutgers.edu

0951-8320/99/$ - see front matter
PII: S0951-8320(98)00081-7

has placed an ''automatic alert tag'' on the aircraft operators of interest.

A performance measure in the SPAS is a pre-defined formula that compares the performance of an aircraft operator to the historical performance of similar aircraft operators, to itself, or to preset limits [2]. A performance measure is not designed to completely describe an aircraft operator. Instead, it is a tool that allows inspectors to evaluate an operator's performance in a specific subject area, as compared to historical patterns. The direction of a performance measure's value (i.e., negative or positive) is intended to be simple. For example, increasing the value of an operator's event rate that is a measure of accidents, incidents, and occurrences is undesirable, but increasing the value of its profit margin is desirable. One of the SPAS databases is the Service Difficulty Reporting (SDR) system. The SDR submission rate is a SPAS performance measure which is an airworthiness or maintenance-related metric. This paper presents alternative methods for trending the SDR submission rate for operators of large commercial aircraft, such as the DC-9 and 737 aircraft.

1.1. Background on service difficulty reports (SDRs)

An SDR provides FAA safety inspectors with information related to abnormal and potentially unsafe mechanical conditions in aircraft components and equipment. Information is provided on equipment inoperability, such as in-service difficulties, malfunctions, and defects. The report assists the inspectors with planning, directing, controlling, and evaluating certain assigned safety and maintenance programs [5]. The system also provides FAA managers and inspectors with a means for measuring the effectiveness of the self-evaluation techniques being employed by certain segments of the civil aviation industry. The completion of an SDR requires careful review of the reported discrepancy and supporting data. An effective evaluation of the extent of the problem and its causes is essential for determining corrective action. If the opportunity exists, the inspector usually reviews prior reports for possible trends, e.g. vendor problems, manufacturer equipment problems, training, and/or procedural problems. However, there are currently no systematic or quantitative techniques used for identifying possible trends. Trending analysis is based on visual inspection of graphical data plots. Approximately 25 000–30 000 SDRs per year are submitted to the FAA.

It should be noted that the SDR database is not a reliability database, as there is no estimation of failure rates, etc. based upon the SDR submission rate. In spite of some limitations with the data collected, it is possible to draw general impressions by examining SDR submission rates and then using this performance measure to provide the rationale for a more detailed examination of aircraft systems or components.

Estimation of the total number of SDRs in a given time interval that a particular airline would be expected to have, adjusting for age of the aircraft, flight time, and landings, could help to identify situations in need of heightened level of surveillance by the FAA's safety inspectors, e.g., if the airline's number of SDRs is far above or below what should be expected. An excessive number of SDRs in a given time period could suggest mechanical, operating, or design problems with certain aircraft. While too few SDRs reported in a given time may not necessarily be problematic, an expert panel of safety inspectors noted that a very low number of SDRs for an airline in a given time period could possibly suggest organizational or management problems, lack of regulatory compliance, airline maintenance cutbacks, or financial or labor problems. Both situations would merit closer scrutiny by FAA safety inspectors.

Fullwood et al. [6] further support the need to develop new trending methods for SDRs. Their study discovered a relationship between equipment operability reported in the SDR and aviation safety as reported in the FAA's Accident Incident Data System (AIDS) that contains reports primarily compiled from National Transportation Safety Board (NTSB) accident investigations. Although equipment problems are not the only factor in aviation accidents, it is the factor reported in the SDR. Fullwood et al. [6] used both a conventional method, in which reporting frequencies are taken from the SDR database and used with an aircraft reliability block diagram model of critical systems to predict aircraft failure, and a shape analysis that uses the magnitude and shape of the SDR distribution compared with the AIDS distribution to predict aircraft failure.

1.2. Previous SDR research and modeling deficiencies

Luxhøj et al. [7] report on SDR trending models for the DC-9 aircraft, and for completeness, these models are reported in this paper. The DC-9 models use a forecast horizon of one year. The data used in this past research investigation included a subset of the SDR database that had been merged with the Aircraft Utilization (ARS) database for the same set of planes. A data grouping method is used to obtain a ''population'' model. Multiple regression and neural network (NN) models were studied and forecasting accuracy for each method was reported. In this study, the original ungrouped data set appeared to be noisy. A ''population concept'' proved to be a very effective modeling technique both for regression analysis and in the construction of NNs for determining strategic safety inspection indicators. While the population concept is constructive for developing models to predict national norms for SDR reporting, there is a loss of information in grouping the data.

In the current research, SDR trending models are developed for both 6-month and monthly planning horizons for the 737 aircraft. The relationship between other aviation safety variables and the SDR number is also investigated. According to the previous research, flying hours, cycles (i.e. the number of takeoffs and landings), age, Airworthiness and Operations Surveillance Results variables are considered

Table 1
Sample of SDR and ARS ''Merged'' DC-9 data [9]

Aircraft model	Serial number[a]	SDR date	Part name	Part location	Part condition	Estimated age	Estimated flight hours	Estimated landings
DC9	333	84-03-22	Skin	E + E Compt	Cracked	17.74	32 619.03	53 999.20
DC9	333	84-03-22	Skin	Aft bag bin	Cracked	17.74	32 619.03	53 999.20
DC9	333	86-07-07	Skin	Fuselage	Cracked	20.03	36 836.23	60 980.56
DC9	444	80-06-20	Skin	Galley door	Cracked	13.24	34 396.44	33 888.77
DC9	444	81-12-01	Skin	FS625	Corroded	14.69	38 160.55	37 597.32
DC9	444	87-05-11	Skin	Rt wheel well	Cracked	20.14	52 299.10	51 527.19
DC9	444	87-05-11	Skin	STA 580-590	Cracked	20.14	52 299.10	51 527.19

[a] Fictitious serial numbers are used owing to confidentiality of data.

to be important independent variables [7]. Variable definitions are provided in the subsequent sections.

This current research uses data from two specific operators with homogenous fleets consisting entirely of 737 aircraft. It is an extension of the previous DC-9 research, but now provides more meaningful information, as the DC-9 aircraft data were composed of numerous operators with heterogenous fleets and differing operating and maintenance policies.

Multiple regression and NN models are the principal two forecasting methods examined. Autoregression, exponential smoothing and moving average time series forecasting techniques are also evaluated. Both NNs and multivariate statistical methods have their advantages and disadvantages, and many of their properties are complementary. NNs are universal approximators and the learned model may be continuously adapted to new data without needing to store previous data. Recurrent networks are naturally well suited for nonlinear dynamic modeling. However, because NNs function essentially as black boxes, their models give limited physical insight into the data. Conversely, linear multivariate statistical methods provide physically interpretable models and the algorithm used for determining the model parameters for large data sets builds the model in a stepwise manner and has guaranteed convergence. However, linear multivariate statistical methods are unable to capture nonlinear behavior, and the models are usually not adapted continuously to new data. Adapting or extending linear multivariate statistical methods to model nonlinear input–output relationships has helped extend their modeling abilities. These nonlinear multivariate methods lie at the interface of neural and statistical methods and combine properties of neural networks and linear multivariate statistical methods. There have been some efforts on revealing the properties and connections between various linear multivariate statistical methods. However, these efforts have focused on only neural or only statistical methods. Consequently, there has been little cross-fertilization between the two fields, and we are still missing a unifying view that brings together both NNs and linear or nonlinear multivariate statistical methods [8].

The paper is organized as follows: first, the SDR trending models for the DC-9 aircraft are presented; second, some regression modeling technical issues are addressed and then the DC-9 NN models are presented; third, the SDR trending models for the 737 aircraft are provided, and then a final section discusses some managerial considerations involved with SDR trending.

2. SDR trending for the DC-9 aircraft

For completeness, this section includes a description of the DC-9 SDR trending models that were previously developed for the whole of the aircraft, for recorded cracking and corrosion cases, and for major structural component groupings.

2.1. Data description for the DC-9 SDR trending models

Our research team was provided with a subset of the SDR database that had been merged with the Aircraft Utilization (ARS) database for the same set of planes. This merged database was created by Battelle (Rice [9]) and consisted of 1308 observations for the DC-9 aircraft for the period April 1974 to March 1990. Table 1 displays sample data. Only the following quantitative data for each plane were available in the merged database:

- Age;
- Estimated Flight Hours;
- Estimated Number of Landings.

Since actual data on flight hours and landings were not reported directly in the SDR database, the estimated flight hours and estimated landings are derived based upon the original delivery date of the plane to the first airline, the date of the ARS data reference, and the SDR date. The equations developed by Battelle for these derived values are reported in Ref. [9] and are presented below:Estimated Flight Hours = [(SDRDate − ServiceDate)/(ARSDate − ServiceDate)] × FHSCUM,Estimated Number of Landings = [(SDRDate − ServiceDate)/(ARSDate − ServiceDate)] × LDGSCUM where SDR Date = date of the SDR report (SDR database); Service Date = original delivery date of the plane to the first airline (ARS database); ARS Date = date of the ARS report (ARS database); FHSCUM = cumulative fuselage flight hours (ARS database); and LDGSCUM = cumulative fuselage landings (ARS database).

Since the ARS date time lagged the SDR date, Rice extrapolated the quantitative ARS data on flight hours and landings to the SDR date. He developed a multiplier by calculating the ratio of (SDR Date − Service Date/ARS Date − Service Date) and then extrapolated the flight hours and landings at the ARS date to the date of the SDR.

2.2. DC-9 SDR multiple regression models

Initially, regression models were created using the 1308 DC-9 observations in their original format, referred to as the 'ungrouped' data. For the ungrouped data, the number of SDRs for each airplane is based on the cumulative number of data records (each record only represents one SDR). When cases with missing data were eliminated, there were a total of 1229 usable data cases. The coefficients of multiple determination, or R^2 values, for these models were very low with the 'best' model having an R^2 value of 0.2448 and a coefficient of variation (CV) of 69.85. The CV reported here is the ratio of the root mean square error of the model to the sample mean of the dependent variable multiplied by 100 and indicates how well the model fits the data. If the model does not fit the data well, then the CV becomes large. It appeared that there was much noise in the data as a plot of the ungrouped data revealed extensive fluctuations.

2.2.1. DC-9 data grouping strategies

In an attempt to create robust SDR prediction models that will provide SDR profiles for a representative DC-9, different data grouping strategies are used. Such an approach was used in Luxhøj and Jones [10], and Luxhøj [11–13] to create large scale logistics models for the U.S. Navy. These ''population'' models were developed to determine both maintenance and system repair/replacement strategies for large groupings of similar equipment based on operating hours, operating environment, failure mode, etc.

Using multiple regression models, data grouping strategies for age, estimated flight hours, and estimated landings are developed based upon a smaller set of averaged data to predict the total expected number of SDRs/year, the number of SDRs/year for cracked cases, and the number of SDRs/year for corrosion cases for the DC-9 aircraft. The ''best'' grouping strategy for each model case is then selected based upon highest R^2 value.

To provide a means for checking the SDR predictions against existing data, the data were partitioned into two different sets based on aircraft serial numbers. The first set was used to build the prediction model and the second set was used to evaluate the prediction model's performance on unfit data. Such an approach is useful for testing prediction model generality [14]. This approach is also used in NN modeling and is analogous to creating a 'training' set of data to build the model and a 'production' set of data to evaluate model performance on new data. These terms are used in this paper to distinguish between the two data sets. The original data were partitioned into mutually exclusive training and production sets by using serial numbers for the different aircraft. Two-thirds of the data were placed into the training set, and one-third into the production set to facilitate cross-validation checks.

After the data have been partitioned into training and production sets, then a grouping strategy is similarly applied to each data set. For example, an age grouping strategy is outlined below:

1. Group the data to create age '*cohorts*' (i.e. groups of 1, 2, 3,…-year-old planes).
2. Calculate the *average* flight hours and number of landings for each *age cohort*.
3. Calculate the average number of SDRs per number of aircraft in each *age cohort*.

Forward stepwise regression is used where variables are added one at a time. Partial correlation coefficients are examined to identify an additional predictor variable that explains both a significant portion and the largest portion of the error remaining from the first regression equation. The forward stepwise procedure selects the 'best' regression model based on highest R^2 from the following list of possible explanatory variables: age, flight hours, number of landings, age^2, flight hours^2, number of landings^2, age × flight hours, age × number of landings, flight hours × number of landings, flight hours/age, and number of landings/age. The default stopping criterion for the F test to determine which variable enters the model uses a significance level of 0.15. In the second stage of our analysis, the best prediction model was chosen on the basis of lowest MSE on the training and production data, since MSE is a better indicator of predictive accuracy. The quadratic terms were considered in an inherently linear model to evaluate any non-linear relationships and the impact of interaction terms was evaluated. The forward stepwise procedure was used to find a prediction equation with an R^2 value close to 1 and to provide an equation that was economical – one that used only a few independent variables.

As a result of the grouping strategy, all interpretations are now with respect to the average number of SDRs per year. In the example above, the dependent variable becomes the average number of SDRs for a representative DC-9 with a 'profile' of estimated flight hours and estimated landings as defined by its associated age cohort. For the grouped data, we now have the number of SDRs for each airplane with respect to an *interval* (i.e., age, flight hours, or landings).

The different structure of the data between grouped and ungrouped records led to structural differences between the regression models and to the use of different explanatory variables.

The grouping procedure resulted in the following:

Model	No. of data records		'Grouped' no. of data records	
	Training	Production	Training	Production
Overall no. of SDRs	805	424	16	14
No. of SDRs (cracking)	572	306	16	16
No. of SDRs (corrosion)	242	127	10	9

A prediction model for the overall number of SDRs per year for a representative DC-9 that uses the 'age' data grouping strategy is given below:Overall no. of SDRs $=$ $(0.00256264 \times \text{agesq}) - (4.038133 \times 10^{-9} \times \text{fhrsq}) + (0.002347 \times \text{fhr/age}) - 4.173934$.

Note that this prediction model makes use of only three independent variables — the age squared (agesq), the flight hours squared (fhrsq), and flight hours/age (fhr/age). The R^2 value is 0.9297 which indicates that this model can explain 92.97% of variability of the expected number of overall SDRs/year about its mean. This model was developed based upon 16 grouped data records that corresponded to aircraft ranging from approximately 8 to 24 years old.

An important point to remember when using this model is that one must have a sufficiently large data sample of DC-9 aircraft in order to compute ''averages'' of estimated landings and flight hours for a specified aircraft age. The more data that one has, the better one can model a representative aircraft using the data grouping strategy as previously discussed.

2.3. Regression modeling adequacy issues

The regression models were examined for multicollinearity, since a high degree of multicollinearity makes the results not generalizable as the parameter estimates in the model may not be stable due to the high variance of the estimated coefficients. Since flying hours, number of landings, and the age of an aircraft are interrelated, multicollinearity is inherent in the independent variables.

Two statistical measures of multicollinearity are the tolerance (TOL) value and the variance inflation factor (VIF) [15]. The TOL value is equal to one minus the proportion of a variable's variance that is explained by the other predictors. A low TOL value indicates a high degree of collinearity. The VIF is the reciprocal of the TOL value, so a high variance inflation factor suggests a high degree of collinearity present in the model. The VIF and TOL measures assume normality and are typically relative measures. A high TOL value (above 0.10) and a low VIF value (below 10) usually suggest a relatively small degree of multicollinearity [15].

While parsimonious regression models were developed by observing the VIF and TOL measures during model building and selection, an attempt was also made to remove multicollinearity by removing the linear trend from the observed variables. Both the dependent and independent variables were transformed by replacing their observed values with their natural logarithms. While this approach was successful in reducing multicollinearity, the resulting regression models all had higher coefficients of variation and lower R^2 values than models without such variable transformations.

There are times in regression modeling when the assumption of constant error variance (i.e. homoscedasticity) may be unreasonable and heteroscedastic error disturbances will occur. When heteroscedasticity is present, ordinary least-squares estimation places more weight on the observations with large error variances than on those with small error variances. The White Test is used in this study to test for heteroscedasticity [14]. In the White Test, the null hypothesis of homoscedasticity is evaluated and the test does not depend critically on normality. The results of the White Test on the data are reported later in this section.

Alternative grouping strategies to 'age' were also examined. Graphical analysis was used to examine the tradeoff of the number of observations versus adjusted R^2 values to determine interval grouping sizes for estimated landings and estimated flight hours. When using the data grouping strategy of estimated landings, the suggested interval size is 4000 landings for the SDR cracking and corrosion cases and 5500 landings for the total number of SDRs. When using the data grouping strategy of estimated flight hours, the suggested interval grouping size is 4000 h. When analyzing the graphs, the goal is to find an interval grouping size that maximizes the adjusted R^2 value, yet results in the use of a reasonable number of observations (i.e., $n \geq 16$ which corresponds to aircrafts ranging from 8 to 24 years old) to facilitate model development. Also, there were upper limits to the interval sizes for landings and flight hours beyond which too few groups resulted. The adjusted R^2 value is used because the number of predictors is changing for each alternative interval size.

As discussed earlier, prediction models were developed using training data and evaluated on production data. Since the goal was to maximize the accuracy of the SDR predictions, the Mean Square Error (MSE) was used for comparative purposes. Although the MSE has some bias, it is an estimator with very low variance.

Table 2 presents the 'best' SDR regression models for the DC-9 aircraft comparing across grouping strategies, predictor variables, and outcome variable. The table also displays the squared partial correlation coefficients that may be used to assess the relative importance of the different independent variables used in the regression models. The VIFs for the overall SDR and corrosion models are acceptable and suggest a relatively small degree of multicollinearity. However, the VIF for the cracking model suggests a moderate degree of collinearity, and this model should be used with caution as the parameter estimates may not be stable. The application of the White Test resulted in the acceptance of the null hypothesis of homoscedasity at the 5% significance level for all three models and suggests that the assumption of constant error variances is reasonable.

Based on an analysis of the 1229 data observations for merged SDR and ARS data, it appears that the data grouping strategy results in SDR prediction models that may be used to predict expected reporting profiles for a representative DC-9. Confidence intervals may be calculated for the expected number of SDRs/year so that a range of values may be reported along with a point estimate. For example, Fig. 1 displays the residuals and confidence limits for the interval (CLI) that includes the variation for both the mean and the error term. In essence, this figure graphically displays the prediction interval for the overall number of SDRs across all airlines for 95% confidence. Such an approach establishes control limits or threshold levels outside of which national SDR advisory warnings would be posted. In order to construct confidence intervals for a particular age group for a given airline, it is necessary to consider the number of aircraft in that age group owned by that airline. SDR prediction models for each airline could be developed by following the same grouping methodology as outlined above, but with the data partitioned by age and airline. Such models were not developed in the DC-9 study, as only 2 of 22 airlines had a sufficient number of data observations by airline.

It appears that ungrouped SDR and ARS data are not useful for prediction purposes. Grouped data strategies show promise in predicting SDR profiles based on the DC-9 analysis. These data grouping strategies generally result in robust models that are useful in developing aircraft population profiles. A plausible reason for the apparent success of the grouping strategy is that computing the average number of SDRs for an interval (i.e. age, flight hours, number of landings) results in the dependent variable becoming approximately normal due to the Central Limit Theorem.

Of the three prediction models, the model to predict the overall expected number of SDRs appears the 'best'. It has the second highest R^2 value (0.9297), a low degree of multicollinearity, and low MSEs on both the training (0.1953) and production (0.9219) data. For the training data, the magnitude of the $\sqrt{MSE}$ is low relative to the sample mean of 2.63 SDRs (i.e. ratio = 0.168). For the production data, the ratio of the $\sqrt{MSE}$ to the sample mean of 2.92 SDRs is higher (i.e., ratio = 0.329). If the $\sqrt{MSE} > 0.33 \times$ sample mean, then normality is not a reasonable assumption and additional distributional information is needed to construct a useful confidence interval. When compared with the overall SDR prediction model, the model for corrosion cases has a higher R^2 value, but there is a degradation of performance on the production data based on MSE. The ratios of the MSEs to the sample means for the training and production data are 0.095 and 0.567, respectively. The model for corrosion was built on the smallest number of averaged data, and this could account for its degraded performance on the production data. The SDR prediction model for cracking cases has the lowest R^2 value, however, it has the best performance on the production data. The ratios of the MSEs to the sample means are 0.064 and 0.092; thus, dispersion of the residuals around the mean is small. When compared with the overall SDR prediction model, the smaller number of averaged observations used in building the model for cracking may account for the lower R^2 value. Stem-and-leaf displays [16] for all models indicate that the shapes of the distributions for the residuals are

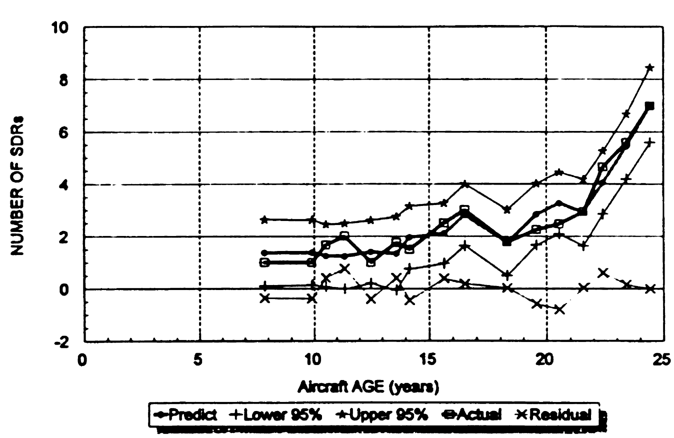


Fig. 1. Residual analysis and 95% confidence limits for overall SDR trending model.

Table 2
DC-9 SDR multiple regression models

Dependent variable	Grouped no. of observations		R^2	CV	MSE		Independent variables	Squared partial correlation coefficient	VIF
	Training data	Production data			Training data	Production data			
Overall no. of SDRs							Agesq	0.9137	11.39
Grouping strategy: age	16	14	0.9297	19.40	0.1953	0.9219	Fhrsq	0.5202	11.19
Increment: 1 year							Fhr/age	0.6070	3.52
No. of SDRs (cracking)							Fhr/age	0.7593	124.81
Grouping strategy: flight hours	16	15	0.7899	6.69	0.0061	0.0265	Age × ldg	0.3764	124.81
Increment: 4000 h									
No. of SDRs (corrosion)							Agesq	0.9661	4.76
Grouping strategy: age	10	9	0.9780	12.29	0.0321	1.6918	Fhrsq	0.9333	9.87
Increment: 1 year							Fhr/age	0.6962	4.32

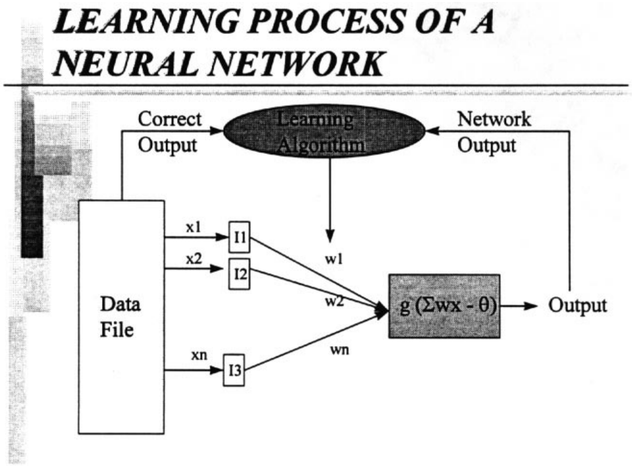


Fig. 2. Neural network learning.

unimodal and bell-shaped. Box plot diagrams of the residuals for all models imply symmetric data sets and that the medians are nearly zero.

A limitation on these prediction models is that the results presented in this paper are based on a relatively small sample of merged DC-9 SDR and ARS data (i.e. 1229 observations) for the period 1974–1990. Generalizing the results to other aircraft types should be done with caution. The value or contribution of this study's findings exists in the methods and techniques used to identify the factors influencing the expected number of SDRs.

2.4. DC-9 SDR neural network models

A parallel research effort focused on the development of NNs to determine patterns in SDR reporting. NNs consist of relatively simple processing elements (nodes or units) connected by links. A unit receives the signal from the input links and computes an activation level that it sends to the next layer along the output links. The computation can be divided into two parts. The first part is a linear function which computes the weighted sum of all the input variables. The second part is a nonlinear function (activation function) which decides whether the output will be greater than the threshold or not [17,18].

Backpropagation is the most popular method in multilayer feedforward networks. It can generalize well on a variety of problems and it is used in most of the existing practical applications of NNs. To develop a backpropagation model, a training set of data patterns that consist of both inputs and the actual outputs observed must be developed. The learning process, as depicted in Fig. 2, is repeated until the error between the actual (correct) and predicted (network) output converges to a predefined threshold. This is accomplished by adjusting the interconnection weights, w_i's, between layers of neurons.

As an alternative to a backpropagation neural network (BPNN), a general regression neural networks (GRNN) is known for its ability to train on sparse data sets. It has been found that a GRNN gives better performance than backpropagation for some problems [19]. It is particularly useful for problems involving continuous function approximation [19]. A GRNN network is a three-layer network that contains one hidden neuron for each training pattern. GRNN training differs from backpropagation networks because training occurs in only one pass. A GRNN is capable of functioning after only a few training patterns have been entered. Both NN architectures were evaluated in this research.

In NN modeling, the R^2 value compares the accuracy of the model with the accuracy of a trivial benchmark model where the prediction is simply the mean of all the sample patterns. A perfect fit would result in an R^2 of 1, a very good fit near 1, and a poor fit near 0. If the NN model predictions are worse than one could predict by just using the mean of the sample case outputs, the R^2 will be 0. Although not precisely interpreted in the same manner as the R^2 in regression modeling, nevertheless, the R^2 from NN modeling can be used as an approximation when comparing model adequacy with a multiple regression model (*NeuroShell 2* [20]).

NN models for SDR prediction were also created using alternative data grouping strategies as previously outlined and the same training and production data sets as those used in the regression analysis. The SDR neural network models are presented in Table 3. Training times for the backpropagation models were insignificant. Since model 'fit' and prediction accuracy were deemed to be most important, the R^2value and MSE were used to select the 'best' NN configuration. The best data grouping strategies as determined from the regression analysis were similarly applied in NN modeling. These NN models may be used to predict

Table 3
DC-9 SDR NN models

Output	No. of patterns		Backpropagation (BP) model[a]			Hybrid model[b]			
	Training data	Production data	R^2	MSE (training)	MSE (production)	R^2	MSE (training)	MSE (production)	n
Overall no. of SDRs	16	14	0.9452	0.152	0.541	0.9603	0.110	2.626	4
No. of SDRs (cracking)	16	15	0.6899	0.009	0.409	0.8404	0.005	0.019	2
No. of SDRs (corrosion)	10	9	0.9411	0.086	3.125	0.9727	0.040	3.502	3

[a] For all BP models, inputs are Age, Fhr and Ldg.
[b] For all hybrid models, inputs are Age, Fhr, Ldg, Class 1,…, Class n, where n is the number of class intervals.

Table 4
Summary of 'best' DC-9 SDR trending models across methods

Modeling method	SDRs overall R^2	MSE		SDRs cracking R^2	MSE		SDRs corrosion R^2	MSE	
	Training data	Training data	Production data	Training data	Training data	Production data	Training data	Training data	Production data
Regression	0.9297	0.1953	0.9219[a]	0.7899	0.0061	0.0265[a]	0.9780	0.0321	1.6918[b]
Backpropagation neural network (BPNN)	0.9452	0.1520	0.5410[b]	0.6899	0.0090	0.4090	0.9411	0.0860	3.1250[a]
BPNN (with predictor variables from regression)	0.9318	0.1890	1.0170	0.6918	0.0090	0.0290	0.9806	0.0290	4.7030
Hybrid neural network	0.9603	0.1100	2.6260	0.8404	0.0050	0.0190[a]	0.9727	0.0400	3.5020

[a] Second choice.
[b] First choice.

the average number of SDRs using a data grouping strategy of one year time increments for the overall number of SDRs and for the number of corrosion cases. To predict the average number of SDRs for cracking cases, the data grouping strategy was based on increments of 4000 flight hours. In all cases, the MSE was lower on the training data than on the production data. Especially note that although the NN for the corrosion case performed well on the training data ($R^2 = 0.9411$, MSE $= 0.086$), the MSE on the production data increased significantly (MSE $= 3.125$). It should also be observed that the model for corrosion cases had the least number of training and production patterns derived from data groupings with the least number of observations of the three models constructed. Thus, this model should be used with caution on unfit data as it does not appear to generalize well.

As in regression modeling, 90% or 95% 'confidence intervals' could be developed for the overall number of SDRs and the number of SDRs for cracking and corrosion cases. These confidence intervals could be displayed in a fashion analogous to quality control charts serving as more refined 'alert' indicators for inspectors that specify upper and lower safety control limits by aircraft type.

The concept of a two-stage hybrid NN is tested in this research to develop SDR prediction models to determine if any incremental improvements could be obtained in prediction accuracy. Table 3 also summarizes the results from these hybrid NNs. The first stage uses a Probabilistic Neural Network (PNN) to classify the age of a DC-9 aircraft into its corresponding '*class*' for the expected number of SDRs. A PNN is a supervised NN that is used to train quickly on sparse data sets [21]. This NN separates input patterns into some defined output categories. In the process of training, the PNN clusters patterns by producing activations in the output layer. The value of the activations correspond to the probability mass function estimate for that category. It was thought that the use of a PNN is this study could be helpful in ''wrinkling'' the SDR data and facilitate the classification of SDRs based upon an input profile of aircraft data.

For the overall SDR prediction model, the PNN is used in this study to classify the number of SDRs into one of four classes, *class 1* for $0 \leq S \leq 2$, *class 2* for $2 \leq S \leq 4$, class 3 for $4 \leq S \leq 6$, and class 4 for $6 \leq S \leq 8$ where S represents the number of SDRs. The PNN is used in the first stage to classify the age of a DC-9 aircraft into its corresponding class for expected number of SDRs. This vector of age and class then is fed into a BPNN to predict the number of SDRs. The second stage then feeds the classified output along with the above quantitative data to a BPNN to predict the number of SDRs. As with multiple regression, models were developed to predict the overall number of SDRs and the number of SDRs for cracking and corrosion cases. For the SDR cracking and corrosion cases, only two and three 'classes' were required, respectively, given the range for the number of SDRs in each case.

In all SDR cases, the prediction results using the hybrid models were better on the training data than from solely using a three layer backpropagation architecture. However, the MSEs using the production data only improved in the cracking case. Further investigations are required with larger data sets to determine the extent of the benefits of a two-stage approach, as the training time significantly increases with the hybrid model.

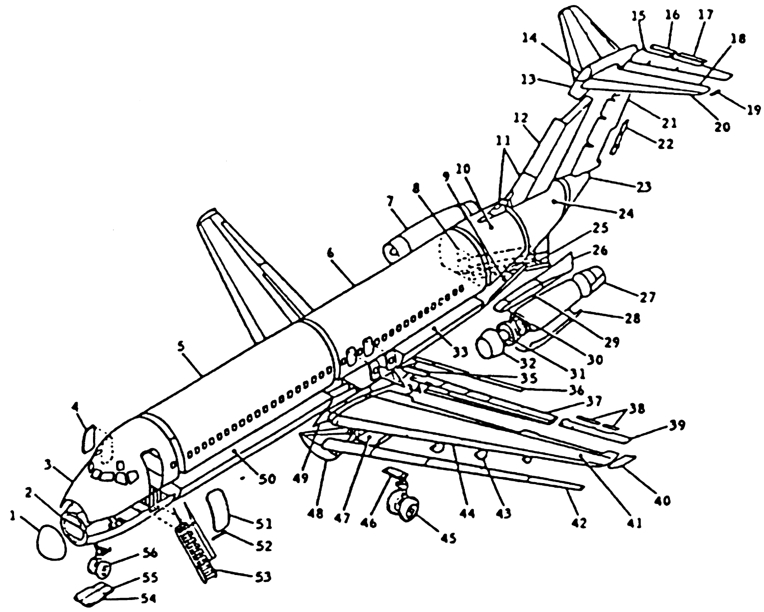
2.4.1. Comparison of DC-9 SDR regression and neural network models

The ''best'' DC-9 SDR trending models compared across modeling methods are identified in Table 4. The 'best' models for each case were selected based upon prediction accuracy with the production data. In the case to predict the overall number of SDRs, the three-layer backpropagation model performs the best. To predict the number of SDR cracking cases, the two-stage hybrid NN is selected, and a regression model is selected as the 'best' method to predict the number of SDR corrosion cases. However, an analysis of Table 4 reveals that the regression models are strong second choices with respect to prediction accuracy. Moreover, the regression models typically take less time to develop than NN models and there is a rich theory for testing regression model adequacy. The modest improvements in predictive accuracy from using a NN in this SDR study do not seem to support the extra 'costs' of computational time and modeling effort required to find a NN that can outperform a regression model.

In the DC-9 study, the information gained from regression analysis regarding the 'best' data grouping strategies helped to improve the performance of a NN model. However, the use of regression analysis to identify the ''best'' set of explanatory variables to use as inputs to a NN needs further investigation.

In this study to develop national SDR prediction models, the original ungrouped data set appeared to be noisy. A 'population concept' proved to be a very effective modeling technique for both regression analysis and in the construction of NNs for determining strategic safety inspection indicators. An important technical issue in using population modeling techniques has to do with failure prediction for parts that have been repaired or replaced. The failure rates of a new part in an old aircraft and new versus repaired parts will affect the inherent characteristics of the aircraft population. One modeling approach is to assume that a repair returns the part to the condition that it was in just prior to failure, so that the part can remain in the same original population. However, this assumption may not be realistic, and repaired parts may need to be modeled with a separate grouping with new population characteristics. For a detailed discussion of preventive maintenance in population models, see Agee and Gallion [22].

While the population concept is constructive for developing models to predict national norms for SDR reporting, there is a loss of information in grouping the data. It is



NO.	Description	NO.	Description	NO.	Description
1.	Radome	20.	Horizontal stabilizer leading edge	39.	Aileron
2.	Fuselage nose lower structure	21.	Rudder	40.	Wing tip
3.	Fuselage nose upper structure	22.	Rudder tab	41.	Wing main structure
4.	Forward service door	23.	Tail cone	42.	Wing slat
5.	Fuselage STA 229 to 588 upper structure	24.	Fuselage tail structure	43.	Flap hinge fairing
6.	Fuselage STA 588 to 996 upper structure	25.	Passenger AFT entrance door stairway	44.	Wing leading edge
7.	Upper cowl door	26.	Pylon AFT panel	45.	Main gear
8.	Passenger AFT entrance stairwell door	27.	Thrust reverser cowling	46.	Main gear outboard door
9.	Fuselage STA 996 to 1087 lower structure	28.	Lower cowl door	47.	Main gear inboard door
10.	Fuselage STA 996 to 1087 upper structure	29.	Pylon center panel	48.	Keel
11.	Dorsal fin	30.	Pylon leading edge	49.	Wing-to-fuselage fillet
12.	Vertical stabilizer	31.	Engine	50.	Fuselage STA 229 to 588 lower structure
13.	Vertical stabilizer tip	32.	Nose cowl	51.	Passenger forward entrance door
14.	Removable tip fairing	33.	Fuselage STA 756 to 996 lower structure	52.	Forward stairwell door
15.	Elevator	34.	Overwing emergency exits	53.	Passenger forward entrance stairwar
16.	Elevator control tab	35.	Flap vane	54.	Forward nose gear doors
17.	Elevator geared tab	36.	Spoiler	55.	AFT nose gear doors
18.	Horizontal stabilizer AFT section	37.	Wing flap	56.	Nose gear
19.	Horizontal stabilizer tip assembly	38.	Aileron tabs		

Fig. 3. Schematic of the DC-9 model 30 aircraft.

recognized that SDR reporting profiles will vary by differences in flying patterns, airlines, location, fleet size, etc. The use of a population concept for two operators of homogeneous fleets is presented in Section 3.

2.5. SDR trending for components of the DC-9 aircraft

In an attempt to further explore the use of NNs to create ‘‘safety alerts’’, Shyur et al. [23] report on the development

Table 5
SDR NN models for components of the DC-9 aircraft [the parameters used were: Learning rate = 0.01; Momentum = 0.05; Initial weight = 0.3; Patterns = 20; Input layers = 3 (Avg. age, Avg. flight hours, Avg. landings); Hidden layers = 21; Output layers = 11 (No. of SDR for each part location)]

Part no.	Part description (.): no. of observations	R^2 value
1	Fuselage nose structure(51)	0.8563
2	Fuselage station 229 to 588(37)	0.8329
3	Fuselage station 588 to 996(81)	0.8504
4	Fuselage station 996 to 1087(68)	0.7382
5	Fuselage tail structure(39)	0.6837
6	Rudder(13)	0.7832
7	Pylon aft panel(7)	0.4926
8	Wing(15)	0.5771
9	Passenger fwd entrance door(27)	0.7948
10	Cargo door(7)	0.8371
11	Aft press blkhd(20)	0.8744

of SDR prediction models for the DC-9 aircraft that use NNs for 11 major structural groupings, such as the cargo door, rudder, wing, nose, etc. The NN models use the three-layer backpropagation learning architecture to predict the expected number of SDRs for cracking cases. A structural schematic of the DC-9 Model 30 aircraft is presented in Fig. 3.

For the 1308 sample data observations, there are only 569 data observations for cracking cases for the DC-9 Model 30 aircraft, and only 390 observations identify the part location. As there were insufficient and incomplete input data for each part location, the part locations were categorized into 11 larger ‘groupings’ as presented in Table 5. Note that the part location numbers in Table 5 do not correspond to the part location numbers in Fig. 3 due to the ‘grouping’ strategy. The corresponding sample sizes for each major part grouping is superscripted in parentheses. Approximately 70% of the cracking cases occurred in the aircraft main fuselage areas and the ‘Fuselage STA 588 to 996’ (recoded as ‘Part 3’) includes 22.2% of the cracking cases.

A three-layer backpropagation architecture is used to classify the SDR cracking cases for data grouped by age in increments of 0.5 years. Moreover, the number of SDRs for one aircraft in a certain age group is calculated. Due to the age ‘grouping’ strategy, only 18 input patterns can be used to train the NN model. The model includes three input neurons (aircraft age, flight hours, and number of landings) and 11 output neurons that identify the number of SDRs in 11 different part locations.

As displayed in Table 5, five of the 11 models have R^2 values above 0.800 which suggests that a BPNN is very effective in predicting the number of SDRs for major structural groupings of part locations. The ‘best’ part location backpropagation models in this study are for AFT Press Bulkhead ($R^2 = 0.8744$), fuselage nose structure ($R^2 = 0.8563$), fuselage stations 588–996 ($R^2 = 0.8504$), cargo door ($R^2 = 0.8371$), and fuselage stations 229–588 ($R^2 = 0.8329$). However, the model cannot predict well in the ‘Pylon AFT panel’ case ($R^2 = 0.4891$), for instance, due to the very limited sample size (i.e. only seven observations). The number of observations for each of the 11 part locations is one major factor that has an influence on the accuracy and efficacy of the model.

3. SDR Trending models for the 737 aircraft

This section presents the results of the current research to develop SDR trending models for specific operators with homogeneous fleets. Two primary data sources are used in this current research: the International Aircraft Operators Information System (IAOIS) and the SPAS. These data sources were not available for use during the development of the DC-9 SDR trending models. The International Aircraft Operators Information System (IAOIS) was originally developed by Wichita State University and is now referred to as the Federal Registry [24]. It is an automated information system which will provide useful aircraft operator information on all United States type-certificated aircraft and airlines worldwide. This system uses commercially available data. The SPAS integrates aircraft operator, aircraft agencies, aircraft type, and aircraft personnel data. The SPAS contains a number of performance measures relating to operations, airworthiness, manuals, records and reporting, maintenance, management, training, and finances.

The SPAS and IAOIS have large amounts of data; however, neither of them is complete. When these two databases are merged we have a relatively complete profile for the aircraft operators. First, an aggregated analysis of the top ten operators of 737 aircraft was attempted. After the query, the top ten operators of 737 aircraft were determined; however, the SPAS operator data only had mixed-model SDRs. Fortunately, Operator A’s and Operator B’s fleets consisted entirely of 737 aircraft. There were 10 351 and 1187 records for Operators A and B, respectively.

Since the SDR numbers in the SPAS database are aggregated for *all* aircraft for a certain operator, the monthly hours and cycles are ‘collapsed’ for all the aircraft to obtain ‘‘average’’ aircraft hours and cycles. For Operator A, 16.5 records form a single data record on average and 145.8 records for Operator B. A new data set was created with the collapsed data for the time period from 30 January 1990 to 30 November 1995 for a total of 71 records.

The following definitions for the SPAS performance measures are used in this research [2]:

AW (Airworthiness) Surveillance Results: Assessment of the results from all airworthiness surveillances performed on an aircraft operator. Indicates the percentage of unfavorable AW surveillance records for a given 1-month period smoothed. AW surveillances focus on the maintenance-related aspects of safety performance (i.e. procedures, log books, equipment, preventive maintenance schedules, etc.)
OPS (Operations) Surveillance Results: Assessment of the results from all OPS surveillances performed on an aircraft

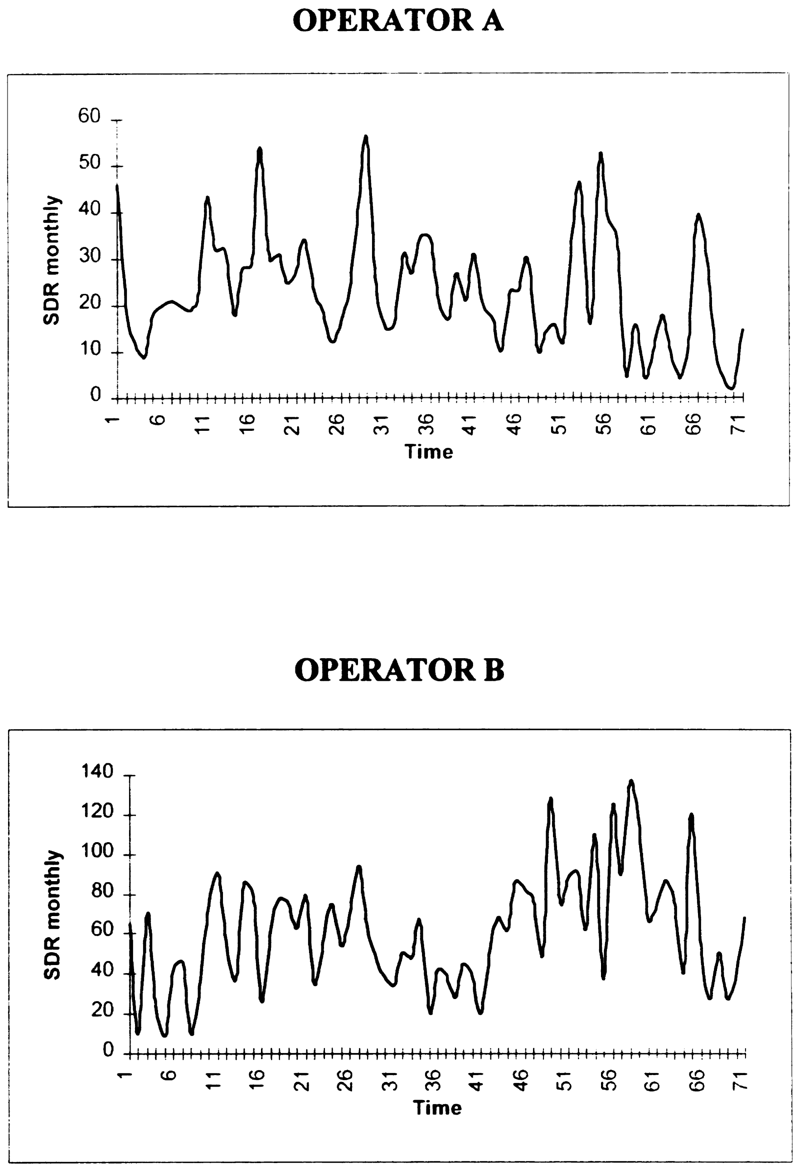


Fig. 4. Actual monthly SDRs for 737 Operators A and B.

operator. Indicates the percentage of unfavorable OPS surveillance records for a given one-month period smoothed. OPS surveillances focus on an operator's procedural knowledge to follow safety regulations (e.g. preflight checks, inflight operations, emergency measures, etc.), and the performance abilities of the crew (e.g. the flight attendants' knowledge of their jobs, the pilot's need to maintain certification, etc.)

SDR FAR Compliance: Assessment of an air operator's compliance with the Federal Aviation Regulations (FARs) for reporting mechanical failures. Indicates the number of SDRs submitted by an aircraft operator over a six-month period.

Sample graphs of monthly SDRs for Operators A and B are exhibited in Fig. 4.

3.1. SDR 737 multiple regression models

We selected the stepwise method since it is the most appropriate method for identifying a model with only a few independent variables. The null hypothesis was evaluated at the 5% level of significance. The null hypothesis here is that a variable has a significant effect on the dependent variable.

The coefficients of multiple determination, or R^2 values, for these first models were not very good, with the 'best' models for Operators A and B having R^2 values of 0.1731 and 0.4573, respectively, and coefficients of variation (CV) of 18.05197 and 21.40486, respectively. As with the DC-9 SDR data, it appeared that there was much noise in the data because a plot of the SDRs revealed extensive fluctuations.

The above results were the best we could determine

Table 6
Classical forecasting results for 737 Operator A

Technique[a]	Dependent variable[b]	MSE for training (80%)	MSE for production (20%)	Variables in the model	Notes
MR	SDR 6-month	107.88	366.45 [3]	SDRAR3, CYCLES	$R^2 = 0.8672$; VIF = 1.027, 1.027
AR	SDR 6-month	117.71	359.30 [2]	LAG1, LAG2	$R^2 = 0.8536$
AR	SDR 6-month	113.23	348.58 [1]	LAG1, LAG2, LAG3	$R^2 = 0.8513$
ES	SDR monthly	159.15	168.84 {1}		$\alpha = 0.3$
AR	SDR monthly	111.48	182.85 {3}	LAG1	$R^2 = 0.0850$
AR	SDR monthly	109.37	181.26 {2}	LAG1, LAG2	$R^2 = 0.1112$

[a] Legends: MR – Multiple Regression; ES – Exponential Smoothing; AR – Autoregression; MA – Moving Average.
[b] Ranks: SDR 6-month [1–best; 3–worst]; SDR monthly {1 – best; 3 – worst}.

without including any lag variables. These initial results were not satisfactory, which indicated that the SDR 6-month moving sum data did not have a linear relationship with the independent variables. From an autocorrelation analysis, we observed that the SDR 6-month moving sum data had a strong relationship with the one-month lag data. The R^2 values reach nearly 0.9. So we included the lag variables [e.g LAG1 (1-month lag), LAG2 (2-month lag), etc.] in our multiple regression models and the results are shown in Table 6.

The multiple regression forecasting model for the SDR 6-month moving sum for Operator A is:Number of SDRs (6-month moving sum) = $-28.849 + 1.037 \times \text{SDRAR3} + 0.0634 \times \text{CYCLES}$ where SDRAR3 = forecast of the number of SDRs from an autoregressive 3 (AR3) time series model and CYCLES = number of aircraft cycles (i.e. number of takeoffs and landings). The AR3 used in this research is of the general form:SDR (AR3) $= a_1\text{SDR}_{t-1} + a_2\text{SDR}_{t-2} + a_3\text{SDR}_{t-3} + b$ where a_1, a_2, and a_3 are coefficients determined from fitting the data, b is a constant, and the subscripts $t-1$, $t-2$, and $t-3$ are used to indicate the number of SDRs from 1 month ago, 2 months ago, etc.

The multiple regression forecasting model for the 6-month moving sum of SDRs for Operator B is:Number of SDRs (6-month moving sum) = $-459.623 + 1.003 \times \text{SDRAR3} + 1.645 \times \text{HRS}$ where SDRAR3 is as defined earlier and HRS = number of aircraft hours. Modeling results are shown in Table 7.

Note that only two variables are included in both forecasting models and the R^2 values for Operators A and B are 0.8672 and 0.8834, respectively, which indicate that these models can explain 86.72% or 88.34% of variability of the expected number of SDRs about their mean. Adjusted R^2 values are 0.8613 and 0.8783, respectively, which are nearly as good as the R^2 values since we have only two independent variables in each model.

The VIFs for both models are acceptable and suggest a small degree of multicollinearity. The application of the White Test resulted in the acceptance of the null hypothesis of homoscedasticity at the 5% significance level for both models and suggests that the assumption of constant error variances is reasonable.

Based on an analysis of the merged data, it appears that a population modeling concept, or data merging strategy, results in SDR prediction models that may be used to predict expected reporting profiles for 737 aircraft of a certain operator. As with the SDR trending models for the DC-9 aircraft, confidence intervals may be calculated for the expected 6-month moving sum of SDRs for the 737 aircraft so that a range of values may be reported along with a point estimate.

Other classical forecasting methods, such as exponential

Table 7
Classical forecasting results for 737 Operator B. (Legend: MR – Multiple Regression; ES – Exponential Smoothing; AR – Autoregression; MA – Moving Average. Ranks: SDR 6-month [1 – best; 3 – worst]; SDR monthly {1 – best; 3 – worst}.

Technique	Dependent variable	MSE for training (80%)	MSE for production (20%)	Variables in the model	Notes
MR	SDR 6-month	990.34	1219.56 [1]	SDRAR3, HRS	$R^2 = 0.8834$; VIF = 1.000447, 1.000447
ES	SDR 6-month	1161.77	1805.65 [3]		$\alpha = 0.01$
AR	SDR 6-month	1113.03	1589.00 [2]	LAG1, LAG2, LAG3	$R^2 = 0.8690$
AR	SDR monthly	600.14	948.61 {2}	LAG1, LAG2	$R^2 = 0.1280$
AR	SDR monthly	557.14	885.65 {1}	LAG1, LAG2, LAG3	$R^2 = 0.2025$
MA	SDR monthly	549.44	1043.70 {3}		5 period

Table 8
Comparison of ''Best'' classical and NN Methods for 737 operators

Method[a]	MSE for training (80%)	MSE for production (20%)
AR – Operator A	113.23	348.58
BPNN – Operator A	160.88	62.35
MR – Operator B	990.34	1219.56
BPNN – Operator B	2607.71	1255.49

[a] AR – Autoregression; BPNN – Backpropagation Neural Network; and MR – Multiple Regression.

smoothing, moving average, and autoregression, were also studied. For Operator A, the two and three order autoregression outperforms multiple regression, but not by a significant amount. For Operator B, multiple regression is the best method followed by a three order autoregression and then exponential smoothing. Modeling details are shown in Tables 6 and 7.

3.2. SDR 737 neural network models

For Operators A and B, other NN structures, such as the GRNN, did not outperform the backpropagation structure. The 'best' models for each operator are listed in Table 8. Graphs of predicted versus actual values for the production data sets for Operators A and B are shown as Figs. 5 and 6, respectively.

Notice that the predicted values fit the actual values very well for the production data for Operator A. The MSE of 62.35 is six-fold lower than that from the classical methods. Conversely, we did not obtain as good a result for Operator B. It is noted that the fleet size for Operator A remains relatively stable at 16 aircraft in the forecast period, while Operator B does not show the same trend as its fleet size is steadily increasing. Even if we include the independent variable SDRs/fleet size, Operator B's NN model still does not provide a good fit to the data. This suggests that an operator's SDR pattern is not related closely to its fleet size. Based on field experience, we have reason to believe that fleet size plays a different role in forecasting SDRs for different operators.

For Operator A, the BPNN model performs better than the classical methods. For Operator B, the MR model performs slightly better than the BPNN model. According to correlation analyses, our data show a low linear relationship between the input and the output variables. Traditional classical methods can only deal with linear relationships but NNs can approximate well on nonlinear relationships.

4. Conclusions and recommendations

Previous research with the development of SDR forecasting models for the DC-9 aircraft were aggregate models that contained mixed fleet data for different operators. The SDR forecasting models in the current research are for a specific-make-model (i.e. 737 aircraft) for two distinct operators

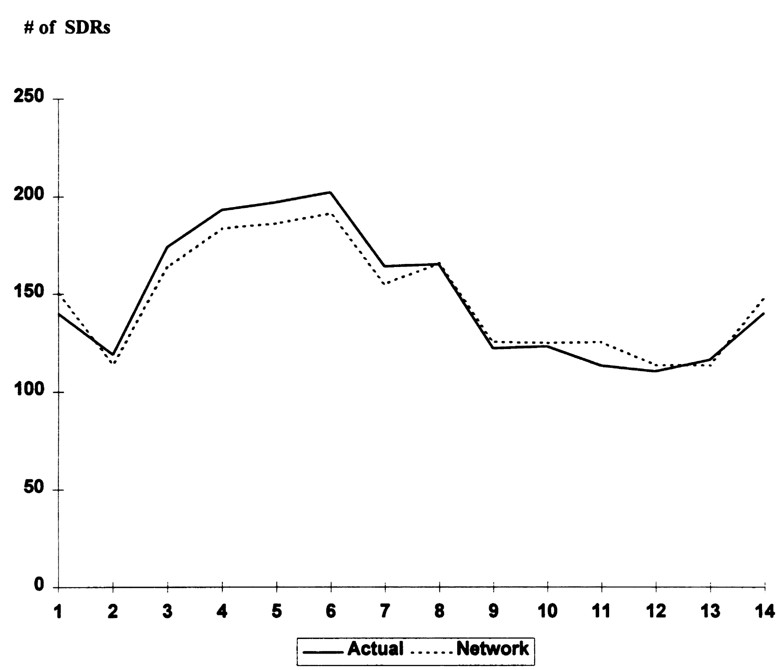


Fig. 5. BPNN results (production data) for 737 Operator A.

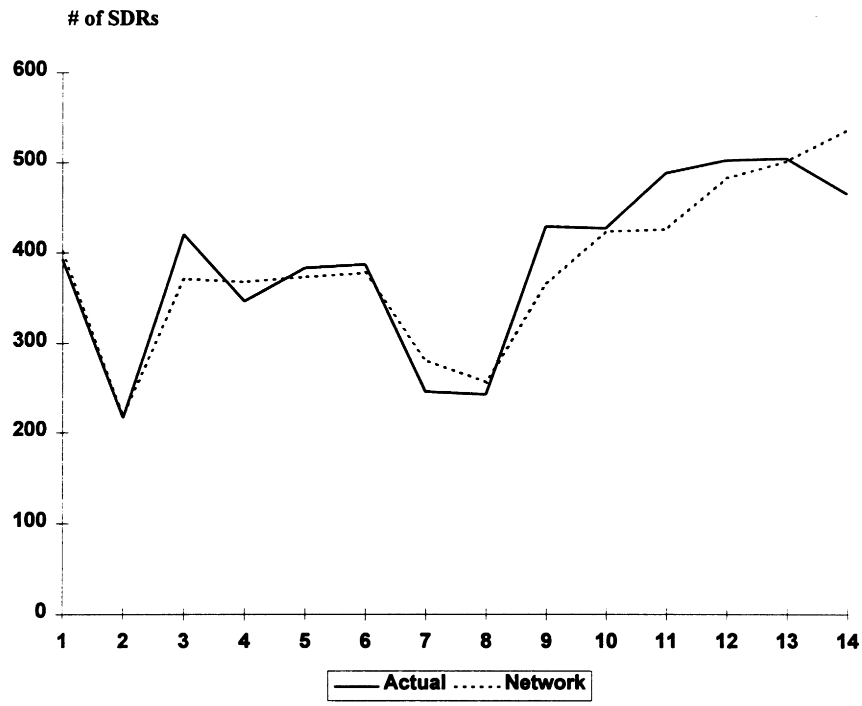


Fig. 6. BPNN results (production data) for 737 Operator B.

which are more meaningful and provide more insight. The monthly number of SDRs was checked as the dependent variable, but poor results were obtained. Models were then developed to trend the 6-month moving sum of SDRs. For Operator A, the NN model is much better than the best classical method. For Operator B, classical methods perform slightly better. It is difficult to generalize from these mixed results, since only two operators whose fleets consisted entirely of 737 aircraft are included in the current study. The NN models were easier to deal with in this study and tended to give better results if the data pattern did not change drastically during the study period. However, classical models were more understandable and explainable.

The 6-month moving SDR sum shows promise in the development of SDR trending models. This study of two aircraft operators suggests that monthly patterns are much harder to trend. Owing to the mixed results obtained from using alternative forecasting methods, this research also suggests that further study is required to ascertain the role that fleet size plays in SDR trending. It should be noted that the SDR submission rate is only one performance measure that focuses solely on equipment inoperability, and that a more comprehensive multivariate data analysis is required in order to develop safety profiles. Nevertheless, the SDR submission rate is an important metric that has aviation safety implications. As such, the forecasting methods presented in this paper offer technical enhancements for SDR trending compared to the current qualitative method of visual observation of graphical plots.

While forecasting the total number of SDRs for a 6-month period is an important first step in planning surveillance activity, the forecasting of SDRs by part type (e.g. landing gear, cargo door, etc.) would enable focused inspections by fleet type on the most problematic aircraft components. The research methods for SDR trending of components for the DC-9 aircraft are being extended to the 737 aircraft.

Acknowledgements

This article is based on research performed at Rutgers University and is supported by grant no. 97-G-005. The contents of this paper reflect the view of the author who is solely responsible for the accuracy of the facts, analyses, conclusions, and recommendations presented herein, and do not necessarily reflect the official view or policy of the Federal Aviation Administration. The author acknowledges the programming assistance of Ms Jiangfeng Cheng.

References

[1] The Federal Aviation Administration (FAA) Plan for research, engineering, and development, volume I: program plan, US Department of Transportation, Report no.100-591, 1989.

[2] Safety Performance Analysis System, SPAS I Version 1.1, Performance measure definitions. Cambridge, MA: US Department of Transportation, Volpe Transportation Systems Center, 1992.

[3] Safety Performance Analysis Subsystem. Cambridge, MA: US Department of Transportation, Volpe National Transportation Systems Center, 1992.

[4] Safety Performance Analysis System II Primer. Cambridge, MA: US Department of Transportation, Volpe Transportation Systems Center, 1997.

[5] Flight Standards Service Difficulty Program. US Department of Transportation, Federal Aviation Administration, 1978.
[6] Fullwood R, Hall R, Martinez G, Uryasev S, Relating aviation service difficulty reports to accident data for safety trend prediction. Upton, NY: Brookhaven National Laboratory Report., #BNL-63018, 1996.
[7] Luxhøj JT, Williams TP, Shyur H. Comparison of regression and neural network models for prediction of inspection profiles for aging aircraft. IIE Transactions on Scheduling and Logistics 1997;29:91–101.
[8] Utojo U, Bakshi B. Connections between neural networks and multivariate statistical methods: an overview. In: Baughman D, Liu Y, editors. Neural Networks in Bioprocessing and Chemical Engineering, New York: Academic Press, 1995.
[9] Rice, R.C., Repair database assessment, Battelle summary report (Contract # DTRS-57- 89-C-00006), March 1991.
[10] Luxhøj JT, Jones MS. A computerized population model for system repair/replacement. Computers and Industrial Engineering 1988;14:345–359.
[11] Luxhøj JT. Sensitivity analysis of maintained systems using a population model: a case study. International Journal of Quality and Reliability Management 1991;8:56–70.
[12] Luxhøj JT. Importance measures for system components in population models. International Journal of Quality and Reliability Management 1991;8:58–69.
[13] Luxhøj JT. Replacement analysis for components of large scale production systems. International Journal of Production Economics 1992;27:97–110.
[14] Pindyck R, Rubinfeld D. Econometric Models and Economic Forecasts. New York: McGraw-Hill, 1991.
[15] Hair J, Anderson R, Tatham R, Black W. Multivariate Data Analysis. 3rd edition. New York: Macmillan Publishing Company, 1992.
[16] Hoaglin DC, Mosteller F, Tukey JW. Understanding Robust and Exploratory Data Analysis. New York: Wiley, 1988.
[17] Russell S, Norvig P. Artificial Intelligence – A Modern Approach. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[18] Simpson P. Artificial Neural Networks. New York: Pergamon Press, 1990.
[19] Specht D. A general regression neural network. IEEE Transactions on Neural Networks 1991;6:568–576.
[20] NeuroShell 2, Ward Systems Group, Frederick, MD, 1993.
[21] Specht D. Probabilistic neural networks. Neural Networks 1990;3:109–118.
[22] Agee MH, Gallion MS, Simulation of population maintenance requirements. In: Algorithmic development and testing of spare parts mortality/support systems software, Vol. II, Research Contract no. N00039-84-C-0346, Naval Electronics System Command, 1986.
[23] Shyur H, Luxhøj JT, Williams TP. Using neural networks to predict component inspection requirements for aging aircraft. Computers and Industrial Engineering 1996;30:257–267.
[24] Hutchinson J, International Aircraft Operator Information System, US Department of Transportation, Federal Aviation Administration, 1996.